DOCUMENT RESUME

ED 206 648 TM 810 538

AUTHOR Day, H. D.; Marshall, Dave

TITLE High Aggregate Stability Coefficients Can Be Obtained for Unstable Traits.

PUB DATE Apr 81

NOTE 21p.; Paper presented at the Annual Meeting of the Southwestern Psychological Association (Houston, TX, April 16-18, 1981).

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS *Behavior Change; Behavior Patterns; *Error of Measurement; Hypothesis Testing; *Reliability; Validity

IDENTIFIERS *Behavioral Stability; Computer Simulation; Epstein (S)

ABSTRACT

In the light of research by Epstein (1979) (which reported that error of measurement in the analysis of behavior stability may be reduced by examining the behavior of aggregate stability coefficients computed for measurements with known stability characteristics), this study examines stability coefficients for computer-generated data sets systematically varying in their stability characteristics. Score for 200 cases across 40 trials were generated to examine the behavior of aggregate stability coefficients for data with known stability characteristics. Forty replications were performed of seven studies, each study varying in the degree of stability characteristics of the scores. Aggregate stability coefficients were computed by correlating the average of even-trial scores with the average of odd-trial scores according to the recommendations of Epstein. Since high stability coefficients were found when only a small percentage of the cases exhibited score stability, the use of aggregate stability coefficients as evidence of stable behavioral dispositions was questioned. (Author/AEF)

# High Aggregate Stability Coefficients Can Be Obtained for Unstable Traits.



H. D. (Jim) Day

Department of Psychology

Texas Woman's University

Box 23029, TWU Station

Denton, Texas 76204

And

Dave Marshall

Department of Mathematics

Texas Woman's University



Paper Presented at the Annual Meeting of the

Southwestern Psychological Association

Houston, Texas

April 16-18, 1981

## Abstract

Scores for 200 cases across 40 trials were generated to examine the behavior of aggregate stability coefficients for data with known stability characteristics. Forty replications were performed of seven studies, each study varying in the degree of stability characteristic of the scores. Aggregate stability coefficients were computed by correlating the average of even-trial scores with the average of odd-trial scores. Since high stability coefficients were found when only a small percentage of the cases exhibited score stability, the use of aggregate stability coefficients as evidence of stable behavioral dispositions was questioned.

# High Aggregate Stability Coefficients Can Be Obtained for Unstable Traits

One of the more enduring questions appearing in the personality literature is whether there are stable behavioral dispositions. Epstein (1979) has recently argued that previous difficulties in demonstrating the stability of behavior can be overcome by employing analyses that reduce the error of measurement. Specifically, Epstein demonstrated that over a series of daily measurements of a behavior, the correlation between the average of even-day measurements and the average of odd-day measurements is generally much higher than the correlation of the measurements for just any two days in the series. According to Epstein's argument, such aggregations of scores reduce the error variance and reveal the underlying behavioral stability.

Since the stability coefficients reported by Epstein (1979) are unusually high (Median = .77 for coefficients in Study 1) and since the application of the method of aggregation to other research problems has been suggested (Epstein, 1980), it seems prudent to examine the behavior of aggregate stability coefficients computed for measurements with known stability characteristics. The present studies, therefore, report stability coefficients for computer-generated data sets systematically varying in their stability characteristics.

## Method

Seven computer simulation studies were conducted, each study involving data randomly produced within different stability constraints. In each study, discrete metric scores ranging from one to 20 were generated for 200 cases and across 40 trials for each case. Aggregate scores for even trials were found by computing the mean for all possible series of even trials beginning with the second trial. Similar aggregates for the odd trials were also computed. Aggregate stability coefficients were then found by correlating the corresponding even-trial and odd-trial aggregates. The first aggregate stability coefficient was, therefore, the correlation between the average of Trials 2 and 4 and the average of Trials 1 and 3; the second aggregate stability score was the correlation between the average of Trials 2, 4, and 6 and the average of Trials 1, 3, and 5; and the last stability coefficient was the correlation between the aggregate of even Trials 2 through 40 and the aggregate of odd Trials 1 through 39. For contrast with the aggregate stability coefficients, a simple reliability coefficient was found by correlating Trial-1 and Trial-2 scores.

In Study 1, data with no stability were generated; all scores were permitted to take on any value between one and 20. Studies 2 through 5 varied stability by holding the scores for a different percentage of the cases constant across the 40 trials while permitting the remaining scores to take on any value between one and 20. Studies 2 through 5 held 1, 5, 12.5, and 25% of the cases' scores constant, respectively. These constant scores, however, were permitted to take on any value between one and 20, but a particular case would retain its constant value across all 40 trials. For example, in Study 2, Case 76 might have a value of 15 for all 40 trials and Case 119 might have a constant value of 11; the remaining 198 cases would have scores randomly varying from one to 20 across the trials. The constant scores of 15 and 11 were each randomly selected from the integer values one to 20, and other replications of Study 2 would have other values for the two constant scores.

In Study 6, half the cases had scores ranging from one to 10 across the 40 trials, and the other 50% of the cases were given scores randomly varying between 11 and 20; there were, therefore, no constant scores in this study. Study 7 also involved no constant scores; half the cases had scores randomly varying over the lower decile (values of one or two), and the other half ranged from three to 20 over the trials. These last two studies were conducted to demonstrate the behavior of the aggregate stability coefficients computed for data produced by the combination of two subsets of scores, each subset having no stability.

## Results and Discussion

Each of the seven studies was replicated 40 times in order that the stability of the aggregate stability coefficients could be determined. The mean, standard error, maximum, and minimum aggregate stability coefficients for the 40 replications of each study are presented in Table 1. By examining the results for Studies 1 through 5 it is apparent that the size of the aggregate stability coefficients varies directly with the degree of stability in the data; mean stability coefficients fluctuate around .00 in Study 1 and become as large as .87 for Study 5. It is also evident from Study 2, however, that significant aggregate stability coefficients can be obtained with as few as 1% of the 200 scores being characterized as stable. Study 3 illustrates that a "quite respectable" aggregate stability coefficient of .50 can be obtained by the present methodology with stability occurring in only 5% of the cases, and with only 25% of the cases exhibiting stability in Study 5, a stability coefficient of .87 was found.

Studies 6 and 7 can be viewed as the combination of two subsets of data having different means but no intra-subset stability. An analogous research setting would involve the combination of two groups of subjects differing in terms of a subject relevant variable, for example, gender. Table 1 indicates that this aggregation of case classifications can produce exceptionally high aggregate stability coefficients.

The results of these seven studies, therefore, reveal several disturbing characteristics of aggregate stability coefficients computed according to the recommendations of Epstein (1979). (1) Significant, and even high, aggregate stability coefficients can be obtained for groups of cases that exhibit rare individual stability. Whereas Epstein (1979) has argued that the success of his stability coefficients is due to the reduction of error of measurement, we suggest that his method of averaging over many trials is so sensitive to case consistencies that it is possible that significant stability coefficients for completely unstable traits in large samples can be obtained when just a few cases exhibit some artifactual stability that may occur because of measurement error. In fact, as the maximum *r*s for Study 1 indicate, for an occasional replication with zero-stable data, the stability coefficient reaches significance. (2) The inclusion of multiple levels of an organismic relevant variable can produce very high stability coefficients when the only stable aspect of the data is group membership. For the cautious interpretation of an aggregate stability coefficient to be possible, therefore, the coefficient should be computed only for samples that are homogeneous with regard to <u>all</u> relevant organismic variables, a requirement that would compromise the utility of a stability measure in practical research applications. (3) The size of the aggregate stability coefficient is directly proportional to the number of trials utilized in its computation. For a coefficient with such a characteristic to be generally useful, each assessment of the stability of a behavior would need to include a sufficient number of measurement of the behavior so that the aggregate coefficient would approach its asymptotic maximum value. (4) For these computer generated data, the size of the aggregate stability coefficient does not seem to be associated in any readily apparent way with the actual degree of stability present in the data. The coefficient of .87 in Study 5, for example, does not seem to logically follow from the fact that 25% of the cases have constant scores.

New statistics for describing behavioral stability and new methodologies for seeking stability will undoubtedly bring the issue of stable behavioral dispositions closer to resolution, and we applaud Professor Epstein's efforts in this regard. The problems revealed by the present simulation studies, however, lead us to endorse only a most cautious application of the method of aggregate stability coefficients.

References

Epstein, S. The stability of behavior: I. On predicting most of the people much of the time. Journal of Personality and Social Psychology, 1979, 37, 1097-1126.

Epstein, S. The stability of behavior: II. Implications for psychological research. American Psychologist, 1980, 35, 790-806.

Table 1

Mean Correlations for Even-day and Odd-day Averages

Number of Averaged Trials

Study 1

| | 1[a] | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\bar{r}$ | .029 | -.001 | .002 | .010 | .014 | .010 | .003 | .004 | -.003 | -.005 |
| $s_r$ | .012 | .013 | .014 | .012 | .011 | .009 | .009 | .009 | .009 | .009 |
| Min | -.133 | -.192 | -.221 | -.217 | -.227 | -.137 | -.154 | -.087 | -.102 | -.128 |
| Max | .198 | .134 | .178 | .125 | .165 | .115 | .140 | .133 | .124 | .118 |
| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| $\bar{r}$ | .007 | .002 | -.001 | -.002 | .004 | .010 | .005 | .009 | .009 | .009 |
| $s_r$ | .008 | .008 | .009 | .009 | .008 | .008 | .008 | .009 | .009 | .010 |
| Min | -.119 | -.140 | -.141 | -.143 | -.118 | -.126 | -.137 | -.140 | -.090 | -.098 |
| Max | .123 | .080 | .098 | .133 | .090 | .097 | .091 | .110 | .121 | .135 |

Study 2

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\bar{r}$ | .020 | .038 | .041 | .049 | .037 | .044 | .064 | .074 | .076 | .075 |
| $s_r$ | .010 | .011 | .011 | .011 | .010 | .010 | .010 | .012 | .011 | .014 |
| Min | -.123 | -.097 | -.065 | -.078 | -.072 | -.052 | -.063 | -.132 | -.107 | -.099 |

Table 1--Continued

Number of Averaged Trials

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Max | .197 | .186 | .230 | .265 | .171 | .190 | .198 | .213 | .220 | .263 |
| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| $\bar{r}$ | .074 | .084 | .099 | .108 | .114 | .124 | .134 | .137 | .150 | .154 |
| $s_r$ | .014 | .014 | .012 | .012 | .013 | .012 | .013 | .014 | .015 | .014 |
| Min | -.145 | -.112 | -.065 | -.071 | -.058 | -.041 | -.047 | -.052 | -.056 | -.047 |
| Max | .266 | .281 | .285 | .320 | .325 | .318 | .357 | .347 | .375 | .356 |

Study 3

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\bar{r}$ | .026 | .073 | .113 | .157 | .196 | .234 | .262 | .293 | .320 | .339 |
| $s_r$ | .011 | .008 | .011 | .011 | .011 | .011 | .012 | .012 | .012 | .012 |
| Min | -.134 | -.071 | -.018 | -.002 | .040 | .033 | .046 | .113 | .132 | .176 |
| Max | .148 | .203 | .326 | .316 | .325 | .390 | .406 | .458 | .475 | .513 |
| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| $\bar{r}$ | .362 | .384 | .397 | .418 | .434 | .450 | .460 | .476 | .488 | .504 |
| $s_r$ | .011 | .012 | .012 | .011 | .011 | .011 | .011 | .011 | .011 | .011 |
| Min | .232 | .248 | .262 | .312 | .293 | .314 | .302 | .305 | .322 | .345 |

Table 1--Continued

Number of Averaged Trials

| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| Max | .534 | .562 | .568 | .575 | .584 | .584 | .604 | .609 | .629 | .654 |

Study 4

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\bar{r}$ | .142 | .240 | .328 | .392 | .445 | .478 | .515 | .550 | .578 | .602 |
| $s_r$ | .010 | .009 | .010 | .010 | .010 | .010 | .009 | .009 | .008 | .008 |
| Min | .005 | .085 | .186 | .236 | .318 | .362 | .401 | .430 | .463 | .494 |
| Max | .268 | .343 | .450 | .513 | .546 | .585 | .610 | .642 | .675 | .681 |
| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| $\bar{r}$ | .624 | .646 | .666 | .680 | .693 | .706 | .722 | .733 | .741 | .753 |
| $s_r$ | .008 | .007 | .007 | .007 | .007 | .006 | .006 | .006 | .006 | .005 |
| Min | .492 | .515 | .526 | .579 | .587 | .618 | .637 | .648 | .666 | .678 |
| Max | .706 | .748 | .773 | .791 | .793 | .795 | .811 | .821 | .812 | .816 |

Study 5

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\bar{r}$ | .271 | .403 | .511 | .571 | .631 | .669 | .701 | .724 | .752 | .770 |
| $s_r$ | .011 | .008 | .007 | .006 | .005 | .005 | .005 | .005 | .004 | .004 |

Table 1--Continued

Number of Averaged Trials

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Min | .073 | .286 | .382 | .439 | .537 | .589 | .640 | .653 | .692 | .722 |
| Max | .399 | .523 | .600 | .646 | .693 | .713 | .767 | .776 | .798 | .811 |
| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| $\bar{r}$ | .786 | .802 | .813 | .823 | .833 | .840 | .848 | .855 | .862 | .866 |
| $s_r$ | .004 | .004 | .004 | .004 | .004 | .004 | .003 | .003 | .003 | .003 |
| Min | .723 | .737 | .749 | .770 | .774 | .783 | .793 | .807 | .815 | .822 |
| Max | .837 | .847 | .858 | .863 | .872 | .882 | .891 | .894 | .901 | .908 |

Study 6

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\bar{r}$ | .647 | .790 | .848 | .880 | .902 | .916 | .927 | .936 | .942 | .948 |
| $s_r$ | .008 | .005 | .003 | .003 | .002 | .002 | .001 | .001 | .001 | .001 |
| Min | .537 | .724 | .799 | .843 | .868 | .893 | .901 | .916 | .922 | .929 |
| Max | .750 | .857 | .900 | .917 | .936 | .944 | .947 | .953 | .961 | .964 |
| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| $\bar{r}$ | .952 | .956 | .959 | .962 | .965 | .967 | .969 | .970 | .972 | .973 |
| $s_r$ | .001 | .001 | .001 | .001 | .001 | .001 | .001 | .001 | .001 | .001 |

Table 1--Continued

Number of Averaged Trials

| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| Min | .932 | .937 | .941 | .944 | .955 | .957 | .960 | .962 | .965 | .966 |
| Max | .965 | .968 | .970 | .971 | .974 | .974 | .977 | .977 | .978 | .980 |

Study 7

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\bar{r}$ | .751 | .857 | .899 | .922 | .935 | .946 | .953 | .959 | .964 | .967 |
| $s_r$ | .003 | .002 | .002 | .001 | .001 | .001 | .000 | .001 | .000 | .000 |
| Min | .709 | .829 | .872 | .908 | .923 | .935 | .948 | .954 | .959 | .963 |
| Max | .789 | .879 | .922 | .940 | .946 | .953 | .958 | .966 | .969 | .973 |
| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| $\bar{r}$ | .970 | .973 | .975 | .977 | .978 | .980 | .981 | .982 | .983 | .984 |
| $s_r$ | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| Min | .966 | .969 | .971 | .972 | .974 | .976 | .978 | .980 | .981 | .981 |
| Max | .975 | .977 | .980 | .981 | .982 | .984 | .984 | .986 | .986 | .986 |

Note: See text for description of studies. $\bar{r}$ = mean r; $s_r$ = standard error of r; Min = minimum r; Max = maximum r. For n = 200, $r \geq .139$ is required for significance at $\alpha = .05$.

[a]Correlation between Trials 1 and 2.